Concentration of Measure And Quantum Entanglement

Zheyuan Wu

February 15, 2026

Contents

Chapter 0: Brief definitions and basic concepts

As the future version of me might forgot everything we have over the summer, as I did for now, I will make a review again from the simple definition to recall the necessary information to tell you why we are here and how we are going to proceed.

This section serve as reference for definitions, notations, and theorems that we will use later. This section can be safely ignored if you are already familiar with the definitions and theorems.

But for the future self who might have no idea what I'm talking about, we will provided detailed definitions to you to understand the concepts.

0.1 Complex vector spaces

The main vector space we are interested in is $\mathbb{C}^n$; therefore, all the linear operators we defined are from $\mathbb{C}^n$ to $\mathbb{C}^n$.

Definition 1. *We denote a vector in vector space as* $|\psi\rangle = (z_1, \ldots, z_n)$ *(might also be infinite dimensional, and* $z_i \in \mathbb{C}$*).*

Here ψ is just a label for the vector, and you don't need to worry about it too much. This is also called the ket, where the counterpart $\langle\psi|$ is called the bra, used to denote the vector dual to ψ; such an element is a linear functional if you really want to know what that is.

Few additional notation will be introduced, in this document, we will follows the notation used in mathematics literature [Axl23]

- $\langle\psi|\varphi\rangle$ is the inner product between two vectors, and $\langle\psi| A |\varphi\rangle$ is the inner product between $A|\varphi\rangle$ and $\langle\psi|$, or equivalently $A^\dagger \langle\psi|$ and $|\varphi\rangle$.
- Given a complex matrix $A = \mathbb{C}^{n\times n}$,
  1. $\overline{A}$ is the complex conjugate of A.

Example

$$A = \begin{bmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \\ 7+i & 8+i & 9+i \end{bmatrix}, \overline{A} = \begin{bmatrix} 1-i & 2-i & 3-i \\ 4-i & 5-i & 6-i \\ 7-i & 8-i & 9-i \end{bmatrix}$$

2. $A^\top$ denotes the transpose of A.

Example

$$A = \begin{bmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \\ 7+i & 8+i & 9+i \end{bmatrix}, A^\top = \begin{bmatrix} 1+i & 4+i & 7+i \\ 2+i & 5+i & 8+i \\ 3+i & 6+i & 9+i \end{bmatrix}$$

3. $A^* = \overline{(A^\top)}$ denotes the complex conjugate transpose, referred to as the adjoint, or Hermitian conjugate of A.

Example

$$A = \begin{bmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \\ 7+i & 8+i & 9+i \end{bmatrix}, A^* = \begin{bmatrix} 1-i & 4-i & 7-i \\ 2-i & 5-i & 8-i \\ 3-i & 6-i & 9-i \end{bmatrix}$$

4. A is unitary if $A^*A = AA^* = I$.
5. A is self-adjoint (hermitian in physics literature) if $A^* = A$.

Motivation of Tensor product

Recall from the traditional notation of product space of two vector spaces V and W, that is, $V \times W$, is the set of all ordered pairs $(|v\rangle, |w\rangle)$ where $|v\rangle \in V$ and $|w\rangle \in W$.

The space has dimension $\dim V + \dim W$.

We want to define a vector space with the notation of multiplication of two vectors from different vector spaces.

That is

$$(|v_1\rangle + |v_2\rangle) \otimes |w\rangle = (|v_1\rangle \otimes |w\rangle) + (|v_2\rangle \otimes |w\rangle)$$
$$|v\rangle \otimes (|w_1\rangle + |w_2\rangle) = (|v\rangle \otimes |w_1\rangle) + (|v\rangle \otimes |w_2\rangle)$$

and enables scalar multiplication by

$$\lambda(|v\rangle \otimes |w\rangle) = (\lambda |v\rangle) \otimes |w\rangle = |v\rangle \otimes (\lambda |w\rangle)$$

And we wish to build a way to associate the basis of V and W with the basis of $V \otimes W$. That makes the tensor product a vector space with dimension $\dim V \times \dim W$.

Definition 2. *Definition of linear functional*

A linear functional is a linear map from V to $\mathbb{F}$.

Note the difference between a linear functional and a linear map.

A generalized linear map is a function $f : V \to W$ satisfying the condition.

- $f(|u\rangle + |v\rangle) = f(|u\rangle) + f(|v\rangle)$
- $f(\lambda |v\rangle) = \lambda f(|v\rangle)$

Definition 3. *A bilinear functional is a bilinear function $\beta : V \times W \to \mathbb{F}$ satisfying the condition that $|v\rangle \to \beta(|v\rangle, |w\rangle)$ is a linear functional for all $|w\rangle \in W$ and $|w\rangle \to \beta(|v\rangle, |w\rangle)$ is a linear functional for all $|v\rangle \in V$.*

The vector space of all bilinear functionals is denoted by $\mathcal{B}(V, W)$.

Definition 4. *Let V, W be two vector spaces.*

Let V' and W' be the dual spaces of V and W, respectively, that is $V' = \{\psi : V \to \mathbb{F}\}$ and $W' = \{\phi : W \to \mathbb{F}\}$, ψ, ϕ are linear functionals.

The tensor product of vectors $v \in V$ and $w \in W$ is the bilinear functional defined by $\forall(\psi, \phi) \in V' \times W'$ given by the notation

$$(v \otimes w)(\psi, \phi) = \psi(v)\phi(w)$$

The tensor product of two vector spaces V and W is the vector space $\mathcal{B}(V', W')$

Notice that the basis of such vector space is the linear combination of the basis of V' and W', that is, if $\{e_i\}$ is the basis of V' and $\{f_j\}$ is the basis of W', then $\{e_i \otimes f_j\}$ is the basis of $\mathcal{B}(V', W')$.

That is, every element of $\mathcal{B}(V', W')$ can be written as a linear combination of the basis.

Since $\{e_i\}$ and $\{f_j\}$ are bases of V' and W', respectively, then we can always find a set of linear functionals $\{\phi_i\}$ and $\{\psi_j\}$ such that $\phi_i(e_j) = \delta_{ij}$ and $\psi_j(f_i) = \delta_{ij}$.

Here $\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$ *is the Kronecker delta.*

$$V \otimes W = \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \phi_i(v) \psi_j(w) : \phi_i \in V', \psi_j \in W' \right\}$$

Note that $\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij} \phi_i(v) \psi_j(w)$ is a bilinear functional that maps $V' \times W'$ to $\mathbb{F}$.

This enables basis-free construction of vector spaces with proper multiplication and scalar multiplication.

Examples of tensor product for vectors

Let $V = \mathbb{C}^2, W = \mathbb{C}^3$, choose bases $\{|0\rangle, |1\rangle\} \subset V, \{|0\rangle, |1\rangle, |2\rangle\} \subset W$.

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = v_1 |0\rangle + v_2 |1\rangle \in V, w = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = w_1 |0\rangle + w_2 |1\rangle + w_3 |2\rangle \in W$$

.

Then the tensor product $v \otimes w$ is given by

$$v \otimes w = \begin{pmatrix} v_1 w_1 & v_1 w_2 & v_1 w_3 \\ v_2 w_1 & v_2 w_2 & v_2 w_3 \end{pmatrix} \in \mathbb{C}^6$$

Examples of tensor product for vector spaces

Let $V = \mathbb{C}^2, W = \mathbb{C}^3$, choose bases $\{|0\rangle, |1\rangle\} \subset V, \{|0\rangle, |1\rangle, |2\rangle\} \subset W$.
Then a basis of the tensor product is

$$\{|00\rangle, |01\rangle, |02\rangle, |10\rangle, |11\rangle, |12\rangle\},$$

where $|ij\rangle := |i\rangle \otimes |j\rangle$.
An example element of $V \otimes W$ is

$$|\psi\rangle = 2\ |0\rangle \otimes |1\rangle + (1+i)\ |1\rangle \otimes |0\rangle - i\ |1\rangle \otimes |2\rangle .$$

With respect to the ordered basis

$$(|00\rangle, |01\rangle, |02\rangle, |10\rangle, |11\rangle, |12\rangle),$$

this tensor corresponds to the coordinate vector

$$|\psi\rangle \longleftrightarrow \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1+i \\ 0 \\ -i \end{pmatrix} \in \mathbb{C}^6.$$

Using the canonical identification

$$\mathbb{C}^2 \otimes \mathbb{C}^3 \cong \mathbb{C}^{2\times 3},$$

where

$$|i\rangle \otimes |j\rangle \longmapsto E_{ij},$$

the same tensor is represented by the matrix

$$|\psi\rangle \longleftrightarrow \begin{pmatrix} 0 & 2 & 0 \\ 1+i & 0 & -i \end{pmatrix}.$$

Definition 5. *The vector space defined by the tensor product is equipped with the unique inner product* $\langle v \otimes w, u \otimes x \rangle_{V \otimes W} : V \otimes W \times V \otimes W \to \mathbb{F}$ *defined by*

$$\langle v \otimes w, u \otimes x \rangle = \langle v, u \rangle_V \langle w, x \rangle_W$$

In practice, we ignore the subscript of the vector space and just write $\langle v \otimes w, u \otimes x \rangle = \langle v, u \rangle \langle w, x \rangle$.
Partial trace

Definition 6. *Let* T *be a linear operator on* $\mathcal{H}$, $(e_1, e_2, \cdots, e_n)$ *be a basis of* $\mathcal{H}$ *and* $(\epsilon_1, \epsilon_2, \cdots, \epsilon_n)$ *be a basis of dual space* $\mathcal{H}^*$. *Then the trace of* T *is defined by*

$$\mathrm{Tr}(T) = \sum_{i=1}^{n} \epsilon_i(T(e_i)) = \sum_{i=1}^{n} \langle e_i, T(e_i)\rangle$$

This is equivalent to the sum of the diagonal elements of T.

Definition 7. *Let T be a linear operator on $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are finite-dimensional Hilbert spaces.*

An operator T on $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$ can be written as

$$T = \sum_{i=1}^{n} a_i A_i \otimes B_i$$

where A_i is a linear operator on $\mathcal{A}$ and B_i is a linear operator on $\mathcal{B}$.

The $\mathcal{B}$-partial trace of T ($\mathrm{Tr}_{\mathcal{B}}(T) : \mathcal{L}(\mathcal{A} \otimes \mathcal{B}) \to \mathcal{L}(\mathcal{A})$) is the linear operator on $\mathcal{A}$ defined by

$$\mathrm{Tr}_{\mathcal{B}}(T) = \sum_{i=1}^{n} a_i \,\mathrm{Tr}(B_i) A_i$$

Or we can define the map $L_v : \mathcal{A} \to \mathcal{A} \otimes \mathcal{B}$ by

$$L_v(u) = u \otimes v$$

Note that $\langle u, L_v^*(u') \otimes v' \rangle = \langle u, u' \rangle \langle v, v' \rangle = \langle u \otimes v, u' \otimes v' \rangle = \langle L_v(u), u' \otimes v' \rangle$.

Therefore, $L_v^* \sum_j u_j \otimes v_j = \sum_j \langle v, v_j \rangle u_j$.

Then the partial trace of T can also be defined by

Let $\{v_j\}$ be a set of orthonormal basis of $\mathcal{B}$.

$$\mathrm{Tr}_{\mathcal{B}}(T) = \sum_j L_{v_j}^*(T) L_{v_j}$$

Definition 8. *Let T be a linear operator on $\mathcal{H} = \mathcal{A} \otimes \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are finite-dimensional Hilbert spaces.*

Let ρ be a state on $\mathcal{B}$ consisting of orthonormal basis $\{v_j\}$ and eigenvalue $\{\lambda_j\}$.

The partial trace of T with respect to ρ is the linear operator on $\mathcal{A}$ defined by

$$\mathrm{Tr}_{\mathcal{A}}(T) = \sum_j \lambda_j L_{v_j}^*(T) L_{v_j}$$

This introduces a new model in mathematics explaining quantum mechanics: the non-commutative probability theory.

0.2 Non-commutative probability theory

The non-commutative probability theory is a branch of generalized probability theory that studies the probability of events in non-commutative algebras.

There are several main components of the generalized probability theory; let's see how we can formulate them, comparing with the classical probability theory.

First, we define the Hilbert space in case one did not make the step from the linear algebra courses like me.

Definition 9. *Hilbert space:*

A Hilbert space is a complete inner product space.

That is, a vector space equipped with an inner product, with the induced metric defined by the norm of the inner product, we have a metric space, which is complete. Reminds that complete mean that every Cauchy sequence, the sequence such that for any $\epsilon > 0$, there exists an N such that for all $m, n \geq N$, we have $|x_m - x_n| < \epsilon$, converges to a limit.

As a side note we will use later, we also defined the Borel measure on a space, here we use the following definition specialized for the space (manifolds) we are interested in.

Definition 10. *Borel measure:*

Let X be a topological space, then a Borel measure $\mu : \mathcal{B}(X) \to [0, \infty]$ on X is a measure on the Borel σ-algebra of X $\mathcal{B}(X)$ satisfying the following properties:

1. *$X \in \mathcal{B}$.*
2. *Close under complement: If $A \subseteq X$, then $\mu(A^c) = \mu(X) - \mu(A)$*
3. *Close under countable unions; If $E_1, E_2, \cdots$ are disjoint sets, then $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$*

In later sections, we will use Lebesgue measure, and Haar measure for various circumstances, their detailed definition may be introduced in later sections.

Example

To introduce an example of Hilbert space we use when studying quantum mechanics, we need to introduce a common inner product used in $\mathbb{C}^n$.

Proposition 11. *The Hermitian inner product on the complex vector space $\mathbb{C}^n$ makes it a Hilbert space.*

Proof. We first verify that the Hermitian inner product

$$\langle u, v \rangle = \sum_{i=1}^{n} \overline{u_i} v_i$$

on $\mathbb{C}^n$ satisfies the axioms of an inner product:

1. **Conjugate symmetry:** For all $u, v \in \mathbb{C}^n$,

$$\langle u, v \rangle = \sum_{i=1}^{n} \overline{u_i} v_i = \overline{\sum_{i=1}^{n} \overline{v_i} u_i} = \overline{\langle v, u \rangle}.$$

2. **Linearity:** For any $u, v, w \in \mathbb{C}^n$ and scalars $a, b \in \mathbb{C}$, we have

$$\langle u, av + bw\rangle = \sum_{i=1}^{n} \overline{u_i}(av_i + bw_i) = a\langle u, v\rangle + b\langle u, w\rangle.$$

3. **Positive definiteness:** For every $u = (u_1, u_2, \cdots, u_n) \in \mathbb{C}^n$, let $u_j = a_j + b_j i$, where $a_j, b_j \in \mathbb{R}$.

$$\langle u, u\rangle = \sum_{j=1}^{n} \overline{u_j} u_j = \sum_{i=1}^{n} (a_i^2 + b_i^2) \geq 0,$$

with equality if and only if $u = 0$.

Therefore, the Hermitian inner product is an inner product.

Next, we show that $\mathbb{C}^n$ is complete with respect to the norm induced by this inner product:

$$\|u\| = \sqrt{\langle u, u\rangle}.$$

Since $\mathbb{C}^n$ is finite-dimensional, every Cauchy sequence (with respect to any norm) converges in $\mathbb{C}^n$. This is a standard result in finite-dimensional normed spaces, which implies that $\mathbb{C}^n$ is indeed complete.

Therefore, since the Hermitian inner product fulfills the inner product axioms and $\mathbb{C}^n$ is complete, the complex vector space $\mathbb{C}^n$ with the Hermitian inner product is a Hilbert space. □

Another classical example of Hilbert space is $L^2(\Omega, \mathcal{F}, P)$, where $(\Omega, \mathcal{F}, P)$ is a measure space (Ω is a set, $\mathcal{F}$ is a σ-algebra on Ω, and P is a measure on $\mathcal{F}$). The L^2 space is the space of all function on Ω that is

1. **square integrable**: square integrable functions are the functions $f : \Omega \to \mathbb{C}$ such that

$$\int_\Omega |f(\omega)|^2 dP(\omega) < \infty$$

with inner product defined by

$$\langle f, g\rangle = \int_\Omega \overline{f(\omega)} g(\omega) dP(\omega)$$

2. **complex-valued**: functions are complex-valued measurable. $f = u + vi$ is complex-valued if u and v are real-valued measurable.

Example

Proposition 12. *$L^2(\Omega, \mathcal{F}, P)$ is a Hilbert space.*

Proof. We check the two conditions of the Hilbert space:

- Completeness: Let (f_n) be a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P)$. Then for any $\epsilon > 0$, there exists an N such that for all $m, n \geq N$, we have

$$\int_\Omega |f_m(\omega) - f_n(\omega)|^2 dP(\omega) < \epsilon^2$$

This means that (f_n) is a Cauchy sequence in the norm of $L^2(\Omega, \mathcal{F}, P)$.

- Inner product: The inner product is defined by

$$\langle f, g\rangle = \int_\Omega \overline{f(\omega)} g(\omega) dP(\omega)$$

This is a well-defined inner product on $L^2(\Omega, \mathcal{F}, P)$. We can check the properties of the inner product:

  - Linearity:

$$\langle af + bg, h\rangle = a\langle f, h\rangle + b\langle g, h\rangle$$

  - Conjugate symmetry:

$$\langle f, g\rangle = \overline{\langle g, f\rangle}$$

  - Positive definiteness:

$$\langle f, f\rangle \geq 0$$

□

Let $\mathcal{H}$ be a Hilbert space. $\mathcal{H}$ consists of complex-valued functions on a finite set $\Omega = \{1, 2, \ldots, n\}$, and the functions $(e_1, e_2, \ldots, e_n)$ form an orthonormal basis of $\mathcal{H}$. (We use Dirac notation $|k\rangle$ to denote the basis vector e_k [Par92].)

As an analog to the classical probability space $(\Omega, \mathcal{F}, \mu)$, which consists of a sample space Ω and a probability measure μ on the state space $\mathcal{F}$, the non-commutative probability space $(\mathcal{H}, \mathcal{P}, \rho)$ consists of a Hilbert space $\mathcal{H}$ and a state ρ on the space of all orthogonal projections $\mathcal{P}$.

The detailed definition of the non-commutative probability space is given below:
Definition 13. *Non-commutative probability space:*

*A non-commutative probability space is a pair $(\mathcal{B}(\mathcal{H}), \mathcal{P})$, where $\mathcal{B}(\mathcal{H})$ is the set of all **bounded** linear operators on $\mathcal{H}$.*

*A linear operator on $\mathcal{H}$ is **bounded** if for all u such that $\|u\| \leq 1$, we have $\|Au\| \leq M$ for some $M > 0$.*

$\mathcal{P}$ is the set of all orthogonal projections on $\mathcal{B}(\mathcal{H})$.

The set $\mathcal{P} = \{P \in \mathcal{B}(\mathcal{H}) : P^* = P = P^2\}$ is the set of all orthogonal projections on $\mathcal{B}(\mathcal{H})$.

Recall from classical probability theory, we call the initial probability distribution for possible outcomes in the classical probability theory as our *state*, simillarly, we need to define the *state* in the non-commutative probability theory.
Definition 14. *Non-commutative probability state:*

Given a non-commutative probability space $(\mathcal{B}(\mathcal{H}), \mathcal{P})$,

A state is a unit vector $\langle\psi|$ in the Hilbert space $\mathcal{H}$, such that $\langle\psi|\,|\psi\rangle = 1$.

Every state uniquely defines a map $\rho : \mathcal{P} \to [0, 1]$, $\rho(P) = \langle\psi|\,P\,|\psi\rangle$ (commonly named as density operator) such that:

- *$\rho(O) = 0$, where O is the zero projection, and $\rho(I) = 1$, where I is the identity projection.*

- *If $P_1, P_2, \ldots, P_n$ are pairwise disjoint orthogonal projections, then $\rho(P_1 + P_2 + \cdots + P_n) = \sum_{i=1}^{n} \rho(P_i)$.*

Note that the pure states are the density operators that can be represented by a unit vector $\langle\psi|$ in the Hilbert space $\mathcal{H}$, whereas mixed states are the density operators that cannot be represented by a unit vector in the Hilbert space $\mathcal{H}$.

If $(|\psi_1\rangle, |\psi_2\rangle, \cdots, |\psi_n\rangle)$ is an orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of ρ, for the eigenvalues $p_1, p_2, \cdots, p_n$, then $p_j \geq 0$ and $\sum_{j=1}^{n} p_j = 1$.

We can write ρ as

$$\rho = \sum_{j=1}^{n} p_j |\psi_j\rangle\langle\psi_j|$$

(Under basis $|\psi_j\rangle$, it is a diagonal matrix with p_j on the diagonal.)

The counterpart of the random variable in the non-commutative probability theory is called an observable, which is a Hermitian operator on $\mathcal{H}$ (for all ψ, ϕ in the domain of A, we have $\langle A\psi, \phi\rangle = \langle\psi, A\phi\rangle$. This kind of operator ensures that our outcome interpreted as probability is a real number).
Definition 15. *Observable:*

Let $\mathcal{B}(\mathbb{R})$ be the set of all Borel sets on $\mathbb{R}$.

An (real-valued) observable (random variable) on the Hilbert space $\mathcal{H}$, denoted by A, is a projection-valued map (measure) $P_A : \mathcal{B}(\mathbb{R}) \to \mathcal{P}(\mathcal{H})$.

Satisfies the following properties:

- *$P_A(\emptyset) = O$ (the zero projection)*
- *$P_A(\mathbb{R}) = I$ (the identity projection)*
- *For any sequence $A_1, A_2, \cdots, A_n \in \mathcal{B}(\mathbb{R})$, the following holds:*
  - *$P_A(\bigcup_{i=1}^{n} A_i) = \bigvee_{i=1}^{n} P_A(A_i)$*
  - *$P_A(\bigcap_{i=1}^{n} A_i) = \bigwedge_{i=1}^{n} P_A(A_i)$*
  - *$P_A(A^c) = I - P_A(A), \forall A \in \mathcal{B}(\mathbb{R})$*

If A is an observable determined by the map $P_A : \mathcal{B}(\mathbb{R}) \to \mathcal{P}(\mathcal{H})$, P_A is a spectral measure (a complete additive orthogonal projection valued measure on $\mathcal{B}(\mathbb{R})$). And every spectral measure can be represented by an observable. [Par05]
Proposition 16. *If A_j are mutually disjoint (that is $P_A(A_i)P_A(A_j) = P_A(A_j)P_A(A_i) = O$ for $i \neq j$), then $P_A(\bigcup_{j=1}^{n} A_j) = \sum_{j=1}^{n} P_A(A_j)$*
Definition 17. *Probability of a random variable:*

Let A be a real-valued observable on a Hilbert space $\mathcal{H}$. ρ be a state. The probability of observing the outcome $E \in \mathcal{B}(\mathbb{R})$ is given by:

$$\mu(E) = \mathrm{Tr}(\rho P_A(E))$$

Restriction of a quantum state to a commutative subalgebra defines an ordinary probability measure.

Example

Let

$$Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The eigenvalues of Z are $+1$ and -1, with corresponding normalized eigenvectors

$$|0\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad |1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The spectral projections are

$$P_Z(\{1\}) = |0\rangle\langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad P_Z(\{-1\}) = |1\rangle\langle 1| = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The associated projection-valued measure P_Z satisfies

$$P_Z(\{1,-1\}) = I, \qquad P_Z(\emptyset) = 0.$$

Let

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The normalized eigenvectors of X are

$$|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right), \qquad |-\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right),$$

with eigenvalues $+1$ and -1, respectively.
The corresponding spectral projections are

$$P_X(\{1\}) = |+\rangle\langle +| = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

$$P_X(\{-1\}) = |-\rangle\langle -| = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Compute

$$P_Z(\{1\})P_X(\{1\}) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \cdot \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

On the other hand,

$$P_X(\{1\})P_Z(\{1\}) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}.$$

Since

$$P_Z(\{1\})P_X(\{1\}) \neq P_X(\{1\})P_Z(\{1\}),$$

the projections do not commute.

Let ρ be a density operator on $\mathbb{C}^2$, i.e.

$$\rho \geq 0, \qquad \mathrm{Tr}(\rho) = 1.$$

For a pure state $|\psi\rangle$, one has

$$\rho = |\psi\rangle \langle\psi| .$$

The probability that a measurement associated with a PVM P yields an outcome in a Borel set $A \in \mathcal{B}$ is

$$\mathbb{P}(A) = \mathrm{Tr}(\rho\, P(A)).$$

For example, let

$$\rho = |0\rangle \langle 0| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then

$$\mathrm{Tr}\big(\rho\, P_Z(\{1\})\big) = 1, \qquad \mathrm{Tr}\big(\rho\, P_X(\{1\})\big) = \frac{1}{2}.$$

Definition 18. *Definition of measurement:*

A measurement (observation) of a system prepared in a given state produces an outcome x, x is a physical event that is a subset of the set of all possible outcomes. For each x, we associate a measurement operator M_x on $\mathcal{H}$.

Given the initial state (pure state, unit vector) u, the probability of measurement outcome x is given by:

$$p(x) = \|M_x u\|^2$$

Note that to make sense of this definition, the collection of measurement operators $\{M_x\}$ must satisfy the completeness requirement:

$$1 = \sum_{x \in X} p(x) = \sum_{x \in X} \|M_x u\|^2 = \sum_{x \in X} \langle M_x u, M_x u\rangle = \langle u, (\sum_{x \in X} M_x^* M_x) u\rangle$$

So $\sum_{x \in X} M_x^* M_x = I$.

Here is Table 1 summarizing the analog of classical probability theory and non-commutative (*quantum*) probability theory [Fer]:

Table 1: Analog of classical probability theory and non-commutative (*quantum*) probability theory

Classical probability	**Non-commutative probability**
Sample space Ω, cardinality $\|\Omega\| = n$, example: $\Omega = \{0, 1\}$	Complex Hilbert space $\mathcal{H}$, dimension $\dim \mathcal{H} = n$, example: $\mathcal{H} = \mathbb{C}^2$
Common algebra of $\mathbb{C}$ valued functions	Algebra of bounded operators $\mathcal{B}(\mathcal{H})$
$f \mapsto \bar{f}$ complex conjugation	$P \mapsto P^*$ adjoint
Events: indicator functions of sets	Projections: space of orthogonal projections $\mathcal{P} \subseteq \mathcal{B}(\mathcal{H})$
functions f such that $f^2 = f = \overline{f}$	orthogonal projections P such that $P^* = P = P^2$
$\mathbb{R}$-valued functions $f = \overline{f}$	self-adjoint operators $A = A^*$
$\mathbb{I}_{f^{-1}(\{\lambda\})}$ is the indicator function of the set $f^{-1}(\{\lambda\})$	$P(\lambda)$ is the orthogonal projection to eigenspace
$f = \sum_{\lambda \in \text{Range}(f)} \lambda \mathbb{I}_{f^{-1}(\{\lambda\})}$	$A = \sum_{\lambda \in \text{sp}(A)} \lambda P(\lambda)$
Probability measure μ on Ω	Density operator ρ on $\mathcal{H}$
Delta measure δ_ω	Pure state $\rho = \|\psi\rangle\langle\psi\|$
μ is non-negative measure and $\sum_{i=1}^n \mu(\{i\}) = 1$	ρ is positive semi-definite and $\text{Tr}(\rho) = 1$
Expected value of random variable f is $\mathbb{E}_\mu(f) = \sum_{i=1}^n f(i)\mu(\{i\})$	Expected value of operator A is $\mathbb{E}_\rho(A) = \text{Tr}(\rho A)$
Variance of random variable f is $\text{Var}_\mu(f) = \sum_{i=1}^n (f(i) - \mathbb{E}_\mu(f))^2 \mu(\{i\})$	Variance of operator A is $\text{Var}_\rho(A) = \text{Tr}(\rho A^2) - \text{Tr}(\rho A)^2$
Covariance of random variables f and g is $\text{Cov}_\mu(f, g) = \sum_{i=1}^n (f(i) - \mathbb{E}_\mu(f))(g(i) - \mathbb{E}_\mu(g))\mu(\{i\})$	Covariance of operators A and B is $\text{Cov}_\rho(A, B) = \text{Tr}(\rho A \circ B) - \text{Tr}(\rho A)\,\text{Tr}(\rho B)$
Composite system is given by Cartesian product of the sample spaces $\Omega_1 \times \Omega_2$	Composite system is given by tensor product of the Hilbert spaces $\mathcal{H}_1 \otimes \mathcal{H}_2$
Product measure $\mu_1 \times \mu_2$ on $\Omega_1 \times \Omega_2$	Tensor product of space $\rho_1 \otimes \rho_2$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$
Marginal distribution $\pi_* v$	Partial trace $\text{Tr}_2(\rho)$

0.3 Quantum physics and terminologies

In this section, we will introduce some terminologies and theorems used in quantum physics that are relevant to our study. Assuming no prior knowledge of quantum physics, we will provide brief definitions and explanations for each term.

One might ask, what is the fundamental difference between a quantum system and a classical system, and why can we not directly apply those theorems in classical computers to a quantum computer? It turns out that quantum error-correcting codes are hard due to the following definitions and features for quantum computing.

Definition 19. *All quantum operations can be constructed by composing four kinds of transformations: (adapted from Chapter 10 of [BZ17])*

1. *Unitary operations. $U(\cdot)$ for any quantum state. It is possible to apply a non-unitary operation for an open quantum system, but that is usually not the focus for quantum computing and usually leads to non-recoverable loss of information that we wish to obtain.*

2. *Extend the system. Given a quantum state $\rho \in \mathcal{H}^N$, we can extend it to a larger quantum system by "entangle" (For this report, you don't need to worry for how quantum entanglement works) it with some new states $\sigma \in \mathcal{H}^K$ (The space where the new state dwells is usually called ancilla system) and get $\rho' = \rho \otimes \sigma \in \mathcal{H}^N \otimes \mathcal{K}$.*

3. *Partial trace. Given a quantum state $\rho \in \mathcal{H}^N$ and some reference state $\sigma \in \mathcal{H}^K$, we can trace out some subsystems and get a new state $\rho' \in \mathcal{H}^{N-K}$.*

4. *Selective measurement. Given a quantum state, we measure it and get a classical bit; unlike the classical case, the measurement is a probabilistic operation. (More specifically, this is some projection to a reference state corresponding to a classical bit output. For this report, you don't need to worry about how such a result is obtained and how the reference state is constructed.)*

$U(n)$ is the group of all $n \times n$ **unitary matrices** over $\mathbb{C}$,

$$U(n) = \{A \in \mathbb{C}^{n \times n} : A^*A = AA^* = I_n\}$$

The uniqueness of such measurement came from the lemma below [Mec]
Lemma 20. *Let $(U(n), \|\cdot\|, \mu)$ be a metric measure space where $\|\cdot\|$ is the Hilbert-Schmidt norm and μ is the measure function.*

The Haar measure on $U(n)$ is the unique probability measure that is invariant under the action of $U(n)$ on itself.

That is, fixing $B \in U(n)$, $\forall A \in U(n)$, $\mu(A \cdot B) = \mu(B \cdot A) = \mu(B)$.

The Haar measure is the unique probability measure that is invariant under the action of $U(n)$ on itself.
Definition 21. *Pure state:*

A random pure state φ is any random variable distributed according to the unitarily invariant probability measure on the pure states $\mathcal{P}(A)$ of the system A, denoted by $\varphi \in_R \mathcal{P}(A)$.

It is trivial that for the space of pure state, we can easily apply the Haar measure as the unitarily invariant probability measure since the space of pure state is S^n for some n. However, for the case of mixed states, that is a bit complicated and we need to use partial tracing to defined the rank-s random states.
Definition 22. *Rank-s random state.*

For a system A and an integer $s \geq 1$, consider the distribution onn the mixed states $\mathcal{S}(A)$ of A induced by the partial trace over the second factor form the uniform distribution on pure states of $A \otimes \mathbb{C}^s$. Any random variable ρ distributed as such will be called a rank-s random states; denoted as $\rho \in_R \mathcal{S}_s(A)$. And $\mathcal{P}(A) = \mathcal{S}_1(A)$.
Proposition 23. *Proposition of indistinguishability:*

Suppose that we have two systems $u_1, u_2 \in \mathcal{H}_1$, the two states are distinguishable if and only if they are orthogonal.

Proof. Ways to distinguish the two states:

1. Set $X = \{0, 1, 2\}$ and $M_i = |u_i\rangle\langle u_i|$, $M_0 = I - M_1 - M_2$
2. Then $\{M_0, M_1, M_2\}$ is a complete collection of measurement operators on $\mathcal{H}$.
3. Suppose the prepared state is u_1, then $p(1) = \|M_1 u_1\|^2 = \|u_1\|^2 = 1$, $p(2) = \|M_2 u_1\|^2 = 0$, $p(0) = \|M_0 u_1\|^2 = 0$.

If they are not orthogonal, then there is no choice of measurement operators to perfectly distinguish the two states.

□

Intuitively, if the two states are not orthogonal, then for any measurement (projection) there exists non-zero probability of getting the same outcome for both states.

0.3.1 Random quantum states

First, we need to define what is a random state in a bipartite system.

Chapter 1

Concentration of Measure And Quantum Entanglement

First, we will build the mathematical model describing the behavior of quantum system and why they makes sense for physicists and meaningful for general publics.

1.1 Motivation

First, we introduce a motivation for introducing non-commutative probability theory to the study of quantum mechanics. This section is mainly based on the book [KM].

1.1.1 Light polarization and the violation of Bell's inequality

The light which comes through a polarizer is polarized in a certain direction. If we fix the first filter and rotate the second filter, we will observe the intensity of the light will change.

The light intensity decreases with α (the angle between the two filters). The light should vanish when $\alpha = \pi/2$.

However, for a system of 3 polarizing filters F_1, F_2, F_3, having directions $\alpha_1, \alpha_2, \alpha_3$, if we put them on the optical bench in pairs, then we will have three random variables P_1, P_2, P_3.
Theorem 24. *Bell's 3 variable inequality:*

For any three random variables P_1, P_2, P_3 in a classical probability space, we have

$$\text{Prob}(P_1 = 1, P_3 = 0) \leq \text{Prob}(P_1 = 1, P_2 = 0) + \text{Prob}(P_2 = 1, P_3 = 0)$$

Proof. By the law of total probability there are only two possibility if we don't observe any light passing the filter pair F_i, F_j, it means the photon is either blocked by F_i or F_j, it means

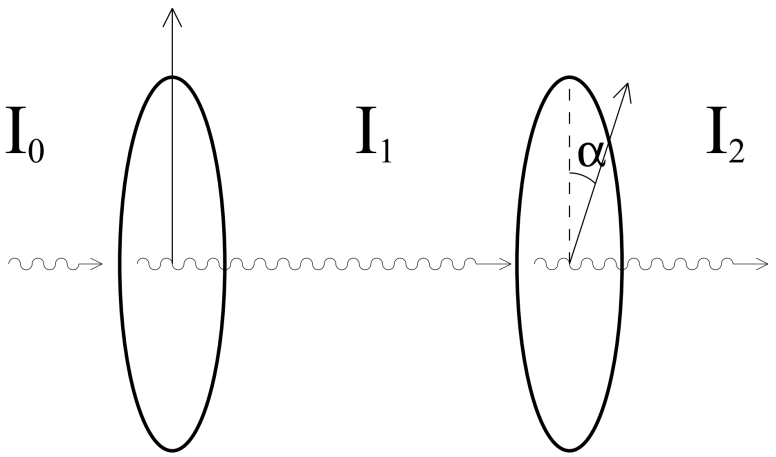

FIG. 1

Figure 1.1: The light polarization experiment, image from [KM]

$$\begin{aligned}\text{Prob}(P_1 = 1, P_3 = 0) &= \text{Prob}(P_1 = 1, P_2 = 0, P_3 = 0) \\ &\quad + \text{Prob}(P_1 = 1, P_2 = 1, P_3 = 0) \\ &\leq \text{Prob}(P_1 = 1, P_2 = 0) + \text{Prob}(P_2 = 1, P_3 = 0)\end{aligned}$$

□

However, according to our experimental measurement, for any pair of polarizers F_i, F_j, by the complement rule, we have

$$\begin{aligned}\text{Prob}(P_i = 1, P_j = 0) &= \text{Prob}(P_i = 1) - \text{Prob}(P_i = 1, P_j = 1) \\ &= \frac{1}{2} - \frac{1}{2}\cos^2(\alpha_i - \alpha_j) \\ &= \frac{1}{2}\sin^2(\alpha_i - \alpha_j)\end{aligned}$$

This leads to a contradiction if we apply the inequality to the experimental data.

$$\frac{1}{2}\sin^2(\alpha_1 - \alpha_3) \leq \frac{1}{2}\sin^2(\alpha_1 - \alpha_2) + \frac{1}{2}\sin^2(\alpha_2 - \alpha_3)$$

If $\alpha_1 = 0, \alpha_2 = \frac{\pi}{6}, \alpha_3 = \frac{\pi}{3}$, then

$$\begin{aligned}\frac{1}{2}\sin^2(-\frac{\pi}{3}) &\leq \frac{1}{2}\sin^2(-\frac{\pi}{6}) + \frac{1}{2}\sin^2(\frac{\pi}{6} - \frac{\pi}{3}) \\ \frac{3}{8} &\leq \frac{1}{8} + \frac{1}{8} \\ \frac{3}{8} &\leq \frac{1}{4}\end{aligned}$$

Other revised experiments (e.g., Aspect's experiment, calcium entangled photon experiment) are also conducted and the inequality is still violated.

1.1.2 The true model of light polarization

The full description of the light polarization is given below:

State of polarization of a photon: $\psi = \alpha|0\rangle + \beta|1\rangle$, where $|0\rangle$ and $|1\rangle$ are the two orthogonal polarization states in $\mathbb{C}^2$.

Polarization filter (generalized 0,1 valued random variable): orthogonal projection P_α on $\mathbb{C}^2$ corresponding to the direction α (operator satisfies $P_\alpha^* = P_\alpha = P_\alpha^2$).

The matrix representation of P_α is given by

$$P_\alpha = \begin{pmatrix} \cos^2(\alpha) & \cos(\alpha)\sin(\alpha) \\ \cos(\alpha)\sin(\alpha) & \sin^2(\alpha) \end{pmatrix}$$

Probability of a photon passing through the filter P_α is given by $\langle P_\alpha\psi, \psi\rangle$; this is $\cos^2(\alpha)$ if we set $\psi = |0\rangle$.

Since the probability of a photon passing through the three filters is not commutative, it is impossible to discuss $\text{Prob}(P_1 = 1, P_3 = 0)$ in the classical setting.

We now show how the experimentally observed probability

$$\frac{1}{2}\sin^2(\alpha_i - \alpha_j)$$

arises from the operator model.

Assume the incoming light is *unpolarized*. It is therefore described by the density matrix

$$\rho = \frac{1}{2}I.$$

Let P_{α_i} and P_{α_j} be the orthogonal projections corresponding to the two polarization filters with angles α_i and α_j.

The probability that a photon passes the first filter P_{α_i} is given by the Born rule:

$$\text{Prob}(P_i = 1) = \text{tr}(\rho P_{\alpha_i}) = \frac{1}{2}\,\text{tr}(P_{\alpha_i}) = \frac{1}{2}$$

If the photon passes the first filter, the post-measurement state is given by the Lüders rule:

$$\rho \longmapsto \rho_i = \frac{P_{\alpha_i}\rho P_{\alpha_i}}{\text{tr}(\rho P_{\alpha_i})} = P_{\alpha_i}.$$

The probability that the photon then passes the second filter is

$$\text{Prob}(P_j = 1 \mid P_i = 1) = \text{tr}(P_{\alpha_i} P_{\alpha_j}) = \cos^2(\alpha_i - \alpha_j).$$

Hence, the probability that the photon passes P_{α_i} and is then blocked by P_{α_j} is

$$\begin{aligned}\text{Prob}(P_i = 1, P_j = 0) &= \text{Prob}(P_i = 1) - \text{Prob}(P_i = 1, P_j = 1) \\ &= \frac{1}{2} - \frac{1}{2}\cos^2(\alpha_i - \alpha_j) \\ &= \frac{1}{2}\sin^2(\alpha_i - \alpha_j).\end{aligned}$$

This agrees with the experimentally observed transmission probabilities, but it should be emphasized that this quantity corresponds to a *sequential measurement* rather than a joint probability in the classical sense.

1.2 Concentration of measure phenomenon

Definition 25. *η-Lipschitz function*

Let (X, dist_X) and (Y, dist_Y) be two metric spaces. A function $f : X \to Y$ is said to be η-Lipschitz if there exists a constant $L \in \mathbb{R}$ such that

$$\text{dist}_Y(f(x), f(y)) \leq L \, \text{dist}_X(x, y)$$

for all $x, y \in X$. And $\eta = \|f\|_{\text{Lip}} = \inf_{L \in \mathbb{R}} L$.

That basically means that the function f should not change the distance between any two pairs of points in X by more than a factor of L.

This is a stronger condition than continuity, every Lipschitz function is continuous, but not every continuous function is Lipschitz.

Lemma 26. *Isoperimetric inequality on the sphere:*

Let $\sigma_n(A)$ denote the normalized area of A on the n-dimensional sphere S^n. That is, $\sigma_n(A) \coloneqq \frac{\text{Area}(A)}{\text{Area}(S^n)}$.

Let $\epsilon > 0$. Then for any subset $A \subset S^n$, given the area $\sigma_n(A)$, the spherical caps minimize the volume of the ϵ-neighborhood of A.

Suppose $\sigma^n(\cdot)$ is the normalized volume measure on the sphere $S^n(1)$, then for any closed subset $\Omega \subset S^n(1)$, we take a metric ball B_Ω of $S^n(1)$ with $\sigma^n(B_\Omega) = \sigma^n(\Omega)$. Then we have

$$\sigma^n(U_r(\Omega)) \geq \sigma^n(U_r(B_\Omega))$$

where $U_r(A) = \{x \in X : d(x, A) < r\}$

Intuitively, the lemma means that the spherical caps are the most efficient way to cover the sphere.

Here, the efficiency is measured by the epsilon-neighborhood of the boundary of the spherical cap.

To prove the lemma, we need to have a good understanding of the Riemannian geometry of the sphere. For now, let's just take the lemma for granted.

1.2.1 Levy's concentration theorem

Theorem 27. *Levy's concentration theorem:*

An arbitrary 1-Lipschitz function $f : S^n \to \mathbb{R}$ concentrates near a single value $a_0 \in \mathbb{R}$ as strongly as the distance function does.

That is,

$$\mu\{x \in S^n : |f(x) - a_0| \geq \epsilon\} < \kappa_n(\epsilon) \leq 2\exp\left(-\frac{(n-1)\epsilon^2}{2}\right)$$

where

$$\kappa_n(\epsilon) = \frac{\int_\epsilon^{\frac{\pi}{2}} \cos^{n-1}(t)dt}{\int_0^{\frac{\pi}{2}} \cos^{n-1}(t)dt}$$

*a_0 is the **Levy mean** of function f, that is, the level set $f^{-1} : \mathbb{R} \to S^n$ divides the sphere into equal halves, characterized by the following equality:*

$$\mu(f^{-1}(-\infty, a_0]) \geq \frac{1}{2} \text{ and } \mu(f^{-1}[a_0, \infty)) \geq \frac{1}{2}$$

We will prove the theorem via the Maxwell-Boltzmann distribution law in this section for simplicity. [Shi14] The theorem will be discussed later in more general cases.

Definition 28. *Gaussian measure:*

We denote the Gaussian measure on $\mathbb{R}^k$ as γ^k.

$$d\gamma^k(x) \coloneqq \frac{1}{\sqrt{2\pi}^k}\exp(-\frac{1}{2}\|x\|^2)dx$$

$x \in \mathbb{R}^k$, $\|x\|^2 = \sum_{i=1}^k x_i^2$ is the Euclidean norm, and dx is the Lebesgue measure on $\mathbb{R}^k$.

Basically, you can consider the Gaussian measure as the normalized Lebesgue measure on $\mathbb{R}^k$ with standard deviation 1.

It also has another name, the Projective limit theorem. [Ver18]

If $X \sim \text{Unif}(S^n(\sqrt{n}))$, then for any fixed unit vector x we have $\langle X, x\rangle \to N(0,1)$ in distribution as $n \to \infty$.

Lemma 29. *Maxwell-Boltzmann distribution law:*

For any natural number k,

$$\frac{d(\pi_{n,k})_*\sigma^n(x)}{dx} \to \frac{d\gamma^k(x)}{dx}$$

where $(\pi_{n,k})_*\sigma^n$ is the push-forward measure of σ^n by $\pi_{n,k}$.

In other words,

$$(\pi_{n,k})_*\sigma^n \to \gamma^k \text{ weakly as } n \to \infty$$

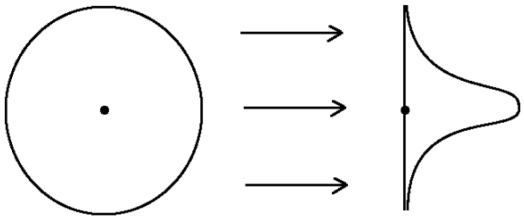

Figure 3.9 The projective central limit theorem: the projection of the uniform distribution on the sphere of radius $\sqrt{n}$ onto a line converges to the normal distribution $N(0,1)$ as $n \to \infty$.

Figure 1.2: Maxwell-Boltzmann distribution law, image from [Ver18]

Proof. We denote the n-dimensional volume measure on $\mathbb{R}^k$ as vol_k.

Observe that $\pi_{n,k}^{-1}(x), x \in \mathbb{R}^k$ is isometric to $S^{n-k}(\sqrt{n - \|x\|^2})$, that is, for any $x \in \mathbb{R}^k$, $\pi_{n,k}^{-1}(x)$ is a sphere with radius $\sqrt{n - \|x\|^2}$ (by the definition of $\pi_{n,k}$).

So,

$$\begin{aligned}\frac{d(\pi_{n,k})_* \sigma^n(x)}{dx} &= \frac{\mathrm{vol}_{n-k}(\pi_{n,k}^{-1}(x))}{\mathrm{vol}_k(S^n(\sqrt{n}))} \\ &= \frac{(n - \|x\|^2)^{\frac{n-k}{2}}}{\int_{\|x\| \le \sqrt{n}} (n - \|x\|^2)^{\frac{n-k}{2}} dx}\end{aligned}$$

as $n \to \infty$.

Note that $\lim_{n\to\infty}(1 - \frac{a}{n})^n = e^{-a}$ for any $a > 0$.

$$(n - \|x\|^2)^{\frac{n-k}{2}} = \left(n(1 - \frac{\|x\|^2}{n})\right)^{\frac{n-k}{2}} \to n^{\frac{n-k}{2}} \exp(-\frac{\|x\|^2}{2})$$

So

$$\begin{aligned}\frac{(n - \|x\|^2)^{\frac{n-k}{2}}}{\int_{\|x\| \le \sqrt{n}} (n - \|x\|^2)^{\frac{n-k}{2}} dx} &= \frac{e^{-\frac{\|x\|^2}{2}}}{\int_{x \in \mathbb{R}^k} e^{-\frac{\|x\|^2}{2}} dx} \\ &= \frac{1}{(2\pi)^{\frac{k}{2}}} e^{-\frac{\|x\|^2}{2}} \\ &= \frac{d\gamma^k(x)}{dx}\end{aligned}$$

□

Now we can prove Levy's concentration theorem, the proof is from [Shi14].

Proof. Let $f_n : S^n(\sqrt{n}) \to \mathbb{R}$, $n = 1, 2, \ldots$, be 1-Lipschitz functions.

Let x and x' be two given real numbers and $\gamma^1(-\infty, x] = \overline{\sigma}_\infty[-\infty, x']$, suppose $\sigma_\infty\{x'\} = 0$, where $\{\sigma_i\}$ is a sequence of Borel probability measures on $\mathbb{R}$.

We want to show that, for all non-negative real numbers ϵ_1 and ϵ_2.

$$\sigma_\infty[x' - \epsilon_1, x' + \epsilon_2] \geq \gamma^1[x - \epsilon_1, x + \epsilon_2]$$

Consider the two spherical cap $\Omega_+ \coloneqq \{f_{n_i} \geq x'\}$ and $\Omega_- \coloneqq \{f_{n_i} \leq x\}$. Note that $\Omega_+ \cup \Omega_- = S^{n_i}(\sqrt{n_i})$.

It is sufficient to show that,

$$U_{\epsilon_1}(\Omega_+) \cup U_{\epsilon_2}(\Omega_-) \subset \{x' - \epsilon_1 \leq f_{n_i} \leq x' + \epsilon_2\}$$

By 1-Lipschitz continuity of f_{n_i}, we have for all $\zeta \in U_{\epsilon_1}(\Omega_+)$, there is a point $\xi \in \Omega_+$ such that $d(\zeta, \xi) \leq \epsilon_1$. So $U_{\epsilon_1}(\Omega_+) \subset \{f_{n_i} \geq x' - \epsilon_1\}$. With the same argument, we have $U_{\epsilon_2}(\Omega_-) \subset \{f_{n_i} \leq x + \epsilon_2\}$.

So the push-forward measure of $(f_{n_i})_*\sigma^{n_i}$ of $[x' - \epsilon_1, x' + \epsilon_2]$ is

$$\begin{aligned}(f_{n_i})_*\sigma^{n_i}[x' - \epsilon_1, x' + \epsilon_2] &= \sigma^{n_i}(x' - \epsilon_1 \leq f_{n_i} \leq x' + \epsilon_2) \\ &\geq \sigma^{n_i}(U_{\epsilon_1}(\Omega_+) \cap U_{\epsilon_2}(\Omega_-)) \\ &= \sigma^{n_i}(U_{\epsilon_1}(\Omega_+)) + \sigma^{n_i}(U_{\epsilon_2}(\Omega_-)) - 1\end{aligned}$$

By the lemma 26, we have

$$\sigma^{n_i}(U_{\epsilon_1}(\Omega_+)) \geq \sigma^{n_i}(U_{\epsilon_1}(B_{\Omega_+})) \quad \text{and} \quad \sigma^{n_i}(U_{\epsilon_2}(\Omega_-)) \geq \sigma^{n_i}(U_{\epsilon_2}(B_{\Omega_-}))$$

By the lemma 29, we have

$$\sigma^{n_i}(U_{\epsilon_1}(\Omega_+)) + \sigma^{n_i}(U_{\epsilon_2}(\Omega_-)) \to \gamma^1[x' - \epsilon_1, x' + \epsilon_2] + \gamma^1[x - \epsilon_1, x + \epsilon_2]$$

Therefore,

$$\begin{aligned}\sigma_\infty[x' - \epsilon_1, x' + \epsilon_2] &\geq \liminf_{i \to \infty} (f_{n_i})_*\sigma^{n_i}[x' - \epsilon_1, x' + \epsilon_2] \\ &\geq \gamma^1[x' - \epsilon_1, \infty) \cap \gamma^1(-\infty, x + \epsilon_2] - 1 \\ &= \gamma^1[x - \epsilon_1, x + \epsilon_2]\end{aligned}$$

□

The full proof of Levy's concentration theorem requires more digestion for cases where $\overline{\sigma}_\infty \neq \delta_{\pm\infty}$ but I don't have enough time to do so. This section may be filled in the next semester.

1.3 The application of the concentration of measure phenomenon in non-commutative probability theory

In quantum communication, we can pass classical bits by sending quantum states. However, by the indistinguishability (Proposition 23) of quantum states, we cannot send an infinite number of classical bits over a single qubit. There exists a bound for zero-error classical communication rate over a quantum channel.

Theorem 30. *Holevo bound:*

The maximal amount of classical information that can be transmitted by a quantum system is given by the Holevo bound. $\log_2(d)$ *is the maximum amount of classical information that can be transmitted by a quantum system with d levels (that is, basically, the number of qubits).*

The proof of the Holevo bound can be found in [NC10]. In current state of the project, this theorem is not heavily used so we will not make annotated proof here.

1.3.1 Quantum communication

To surpass the Holevo bound, we need to use the entanglement of quantum states.

Definition 31. *Bell state:*

The Bell states are the following four states:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle), \quad |\Phi^-\rangle = \frac{1}{\sqrt{2}}(|00\rangle - |11\rangle)$$

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle), \quad |\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$$

These are a basis of the 2-qubit Hilbert space.

1.3.2 Superdense coding and entanglement

The description of the superdense coding can be found in [GMS15] and [Hay10].

Suppose A and B share a Bell state (or other maximally entangled state) $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, where A holds the first part and B holds the second part.

A wishes to send 2 **classical bits** to B.

A performs one of four Pauli unitaries (some fancy quantum gates named X, Y, Z, I) on the combined state of entangled qubits $\otimes$ one qubit. Then A sends the resulting one qubit to B.

This operation extends the initial one entangled qubit to a system of one of four orthogonal Bell states.

B performs a measurement on the combined state of the one qubit and the entangled qubits he holds.

B decodes the result and obtains the 2 classical bits sent by A.

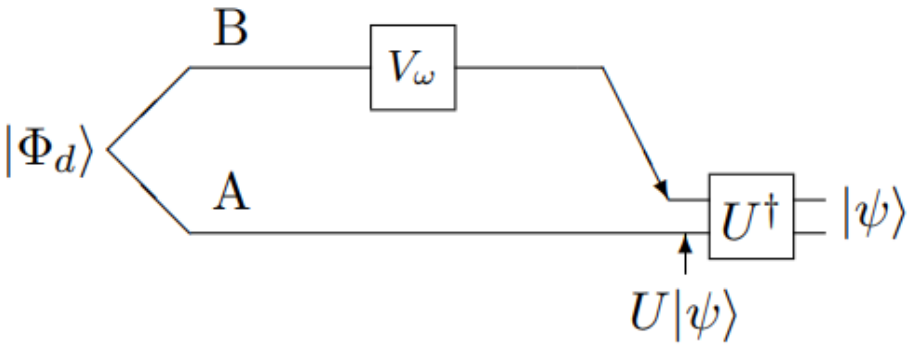

Figure 1.3: Superdense coding, image from [Hay10]

Note that superdense coding is a way to send 2 classical bits of information by sending 1 qubit with 1 entangled qubit. **The role of the entangled qubit** is to help them to distinguish the 4 possible states of the total 3 qubits system where 2 of them (the pair of entangled qubits) are mathematically the same.

Additionally, no information can be gained by measuring a pair of entangled qubits. To send information from A to B, we need to physically send the qubits from A to B. That means, we cannot send information faster than the speed of light.

1.3.3 Hayden's concentration of measure phenomenon

The application of the concentration of measure phenomenon in the superdense coding can be realized in random sampling the entangled qubits [Hay10]:

It is a theorem connecting the following mathematical structure:

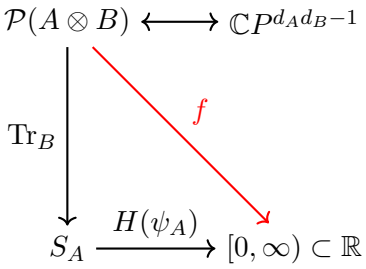

Figure 1.4: Mathematical structure for Hayden's concentration of measure phenomenon

- The red arrow is the concentration of measure effect. $f = H(\mathrm{Tr}_B(\psi))$.
- S_A denotes the mixed states on A.

To prove the concentration of measure phenomenon, we need to analyze the following elements involved in figure 1.4:

The existence and uniqueness of the Haar measure is a theorem in compact lie group theory. For this research topic, we will not prove it.

Due to time constrains of the projects, the following lemma is demonstrated but not investigated thoroughly through the research:

Lemma 32. *Page's lemma for expected entropy of mixed states*

Choose a random pure state $\sigma = |\psi\rangle\langle\psi|$ from $A' \otimes A$.

The expected value of the entropy of entanglement is known and satisfies a concentration inequality known as Page's formula [Pag; San95; BZ17][15.72].

$$\mathbb{E}[H(\psi_A)] = \frac{1}{\ln(2)} \left(\sum_{j=d_B+1}^{d_A d_B} \frac{1}{j} - \frac{d_A - 1}{2d_B} \right) \geq \log_2(d_A) - \frac{1}{2\ln(2)} \frac{d_A}{d_B}$$

It basically provides a lower bound for the expected entropy of entanglement. Experimentally, we can have the following result (see Figure 1.5):

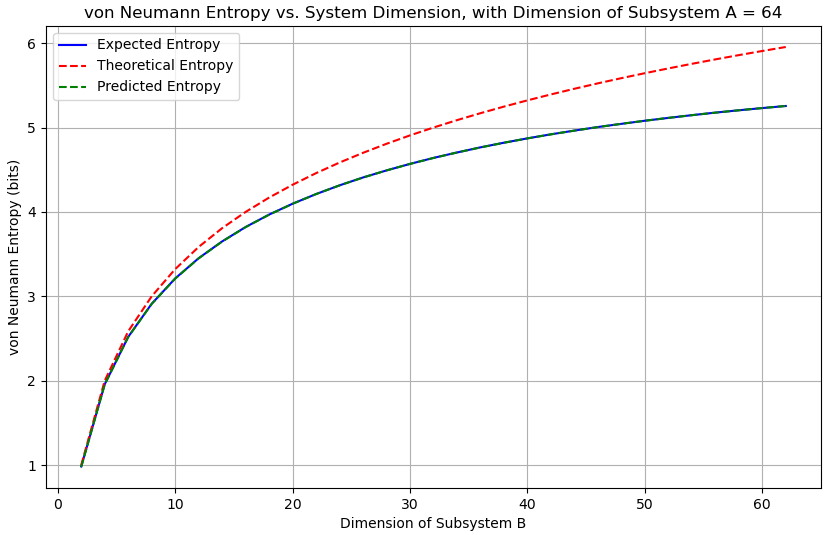

Figure 1.5: Entropy vs dimension

Then we have bound for Lipschitz constant η of the map $S(\varphi_A) : \mathcal{P}(A \otimes B) \to \mathbb{R}$

Lemma 33. *The Lipschitz constant η of $S(\varphi_A)$ is upper bounded by $\sqrt{8}\log_2(d_A)$ for $d_A \geq 3$.*

Proof. Consider the Lipschitz constant of the function $g : A \otimes B \to \mathbb{R}$ defined as $g(\varphi) = H(M(\varphi_A))$, where $M : A \otimes B \to \mathcal{P}(A)$ is any fixed complete von Neumann measurement and $H : \mathcal{P}(A) \otimes \mathcal{P}(B) \to \mathbb{R}$ is the Shannon entropy.

Let $\{|e_j\rangle_A\}$ be the orthonormal basis for A and $\{|f_k\rangle_B\}$ be the orthonormal basis for B. Then we decompose the state as spectral form $|\varphi\rangle = \sum_{j=1}^{d_A} \sum_{k=1}^{d_B} \varphi_{jk} |e_j\rangle_A |f_k\rangle_B$.

By unitary invariance, suppose $M_j = |e_j\rangle \langle e_j|_A$, and define

$$p_j(\varphi) = \langle e_j| \varphi_A |e_j\rangle = \sum_{k=1}^{d_B} |\varphi_{jk}|^2$$

Then

$$g(\varphi) = H(M(\varphi_A)) = -\sum_{j=1}^{d_A} p_j(\varphi) \log_2(p_j(\varphi))$$

Let $h(p) = -p\log_2(p)$, $h(p) = -\frac{p\ln p}{\ln 2}$, and $h'(p) = -\frac{\ln p+1}{\ln 2}$. Let $\varphi_{jk} = x_{jk} + iy_{jk}$, then $p_j(\varphi) = \sum_{k=1}^{d_B}(x_{jk}^2 + y_{jk}^2)$, $\frac{\partial p_j}{\partial x_{jk}} = 2x_{jk}$, $\frac{\partial p_j}{\partial y_{jk}} = 2y_{jk}$.

Therefore

$$\frac{\partial g}{\partial x_{jk}} = \frac{\partial g}{\partial p_j}\frac{\partial p_j}{x_{jk}} = -\frac{1+\ln p_j}{\ln 2}\cdot 2x_{jk} \qquad \frac{\partial g}{\partial y_{jk}} = -\frac{1+\ln p_j}{\ln 2}\cdot 2y_{jk}$$

Then the lipschitz constant of g is

$$\begin{aligned}
\eta^2 &= \sup_{\langle\varphi|\varphi\rangle\le 1} \nabla g \cdot \nabla g \\
&= \sum_{j=1}^{d_A}\sum_{k=1}^{d_B} \left(\frac{\partial g}{\partial x_{jk}}\right)^2 + \left(\frac{\partial g}{\partial y_{jk}}\right)^2 \\
&= \sum_{j=1}^{d_A}\sum_{k=1}^{d_B} \frac{4(x_{jk}^2 + y_{jk}^2)}{(\ln 2)^2}[1+\ln p_j(\varphi)]^2 \\
&= \sum_{j=1}^{d_A}\sum_{k=1}^{d_B} \frac{4|\varphi_{jk}|^2}{(\ln 2)^2}[1+\ln p_j(\varphi)]^2
\end{aligned}$$

Note that $\sum_{k=1}^{d_B} |\varphi_{jk}|^2 = p_j(\varphi)$, $\nabla g \cdot \nabla g = \frac{4}{(\ln 2)^2}\sum_{j=1}^{d_A} p_j(\varphi)(1+\ln p_j(\varphi))^2$.

Since $0 \le p_j \le 1$, we have $\ln p_j(\varphi) \le 0$, hence $\sum_{j=0}^{d_A} p_j(\varphi)\ln p_j(\varphi) \le 0$.

$$\begin{aligned}
\sum_{j=1}^{d_A} p_j(\varphi)(1+\ln p_j(\varphi))^2 &= \sum_{j=1}^{d_A} p_j(\varphi)(1 + 2\ln p_j(\varphi) + (\ln p_j(\varphi))^2) \\
&= 1 + 2\sum_{j=1}^{d_A} p_j(\varphi)\ln p_j(\varphi) + \sum_{j=1}^{d_A} p_j(\varphi)(\ln p_j(\varphi))^2 \\
&\le 1 + \sum_{j=1}^{d_A} p_j(\varphi)(\ln p_j(\varphi))^2
\end{aligned}$$

Thus,

$$\begin{aligned}\nabla g \cdot \nabla g &\leq \frac{4}{(\ln 2)^2}[1 + \sum_{j=1}^{d_A} p_j(\varphi)(\ln p_j(\varphi))^2] \\ &\leq \frac{4}{(\ln 2)^2}[1 + (\ln d_A)^2] \\ &\leq 8(\log_2 d_A)^2\end{aligned}$$

Proving $\sum_j^{d_A} p_j(\varphi) \ln p_j(\varphi) \leq (\ln d_A)^2$ for $d_A \geq 3$ takes some efforts and we will continue that later.

Consider any two unit vectors $|\varphi\rangle$ and $|\psi\rangle$, assume $S(\varphi_A) \leq S(\psi_A)$. If we choose the measurement M to be along the eigenbasis of φ_A, $H(M(\varphi_A)) = S(\varphi_A)$ and we have

$$S(\psi_A) - S(\varphi_A) \leq H(M(\psi_A)) - H(M(\varphi_A)) \leq \eta \| \, |\psi\rangle - |\varphi\rangle \, \|$$

Thus the lipschitz constant of $S(\varphi_A)$ is upper bounded by $\sqrt{8}\log_2(d_A)$. □

From Levy's lemma, we have

If we define $\beta = \frac{1}{\ln(2)} \frac{d_A}{d_B}$, then we have

$$\Pr[H(\psi_A) < \log_2(d_A) - \alpha - \beta] \leq \exp\left(-\frac{1}{8\pi^2 \ln(2)} \frac{(d_A d_B - 1)\alpha^2}{(\log_2(d_A))^2}\right)$$

where $d_B \geq d_A \geq 3$ [HLW06].

Experimentally, we can have the following result:

As the dimension of the Hilbert space increases, the chance of getting an almost maximally entangled state increases (see Figure 1.6).

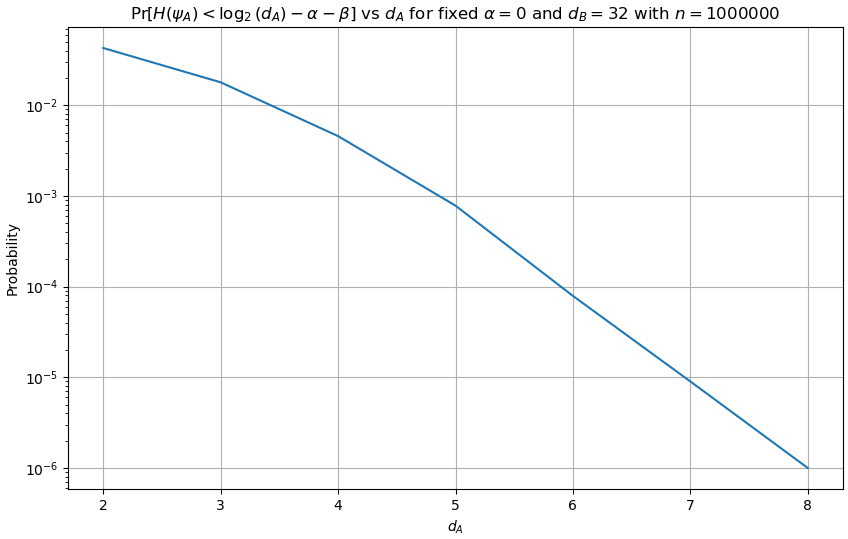

Figure 1.6: Entropy vs d_A

Chapter 2

Levy's family and observable diameters

In this section, we will explore how the results from Hayden's concentration of measure theorem can be understood in terms of observable diameters from Gromov's perspective and what properties it reveals for entropy functions.

We will try to use the results from previous sections to estimate the observable diameter for complex projective spaces.

2.1 Observable diameters

Recall from previous sections, an arbitrary 1-Lipschitz function $f : S^n \to \mathbb{R}$ concentrates near a single value $a_0 \in \mathbb{R}$ as strongly as the distance function does.

Definition 34. *Let X be a topological space with the following:*

1. *X is a complete (every Cauchy sequence converges)*
2. *X is a metric space with metric d_X*
3. *X has a Borel probability measure μ_X*

*Then (X, d_X, μ_X) is a **metric measure space**.*

Definition 35. *Let (X, d_X) be a metric space. The **diameter** of a set $A \subset X$ is defined as*

$$\operatorname{diam}(A) = \sup_{x,y \in A} d_X(x, y).$$

Definition 36. *Let (X, d_X, μ_X) be a metric measure space, For any real number $\alpha \leq 1$, the **partial diameter** of X is defined as*

$$\operatorname{diam}(A; \alpha) = \inf_{A \subseteq X | \mu_X(A) \geq \alpha} \operatorname{diam}(A).$$

This definition generalize the relation between the measure and metric in the metric-measure space. Intuitively, the space with smaller partial diameter can take more volume with the same diameter constrains.

However, in higher dimensions, the volume may tend to concentrates more around a small neighborhood of the set, as we see in previous chapters with high dimensional sphere as example. We can safely cut $\kappa > 0$ volume to significantly reduce the diameter, this yields better measure for concentration for shapes in spaces with high dimension.

Definition 37. *Let X be a metric-measure space, Y be a metric space, and $f : X \to Y$ be a 1-Lipschitz function. Then $f_*\mu_X = \mu_Y$ is a push forward measure on Y.*

*For any real number $\kappa > 0$, the κ-**observable diameter with screen** Y is defined as*

$$\mathrm{ObserDiam}_Y(X;\kappa) = \sup\{\mathrm{diam}(f_*\mu_X; 1-\kappa)\}$$

*And the **obbservable diameter with screen** Y is defined as*

$$\mathrm{ObserDiam}_Y(X) = \inf_{\kappa>0} \max\{\mathrm{ObserDiam}_Y(X;\kappa)\}$$

*If $Y = \mathbb{R}$, we call it the **observable diameter**.*

If we collapse it naively via

$$\inf_{\kappa>0} \mathrm{ObserDiam}_Y(X;\kappa),$$

we typically get something degenerate: as $\kappa \to 1$, the condition "mass $\geq 1-\kappa$" becomes almost empty space, so $\mathrm{diam}(\nu; 1-\kappa)$ can be forced to be 0 (take a tiny set of positive mass), and hence the infimum tends to 0 for essentially any non-atomic space.

This is why one either:

1. keeps $\mathrm{ObserDiam}_Y(X;\kappa)$ as a *function of* κ (picking κ to be small but not 0), or
2. if one insists on a single number, balances "spread" against "exceptional mass" by defining $\mathrm{ObserDiam}_Y(X) = \inf_{\kappa>0} \max\{\mathrm{ObserDiam}_Y(X;\kappa), \kappa\}$ as above.

The point of the $\max\{\cdot, \kappa\}$ is that it prevents cheating by taking κ close to 1: if κ is large then the maximum is large regardless of how small $\mathrm{ObserDiam}_Y(X;\kappa)$ is, so the infimum is forced to occur where the exceptional mass and the observable spread are small.

Few additional proposition in [Shi14] will help us to estimate the observable diameter for complex projective spaces.

Proposition 38. *Let X and Y be two metric-measure spaces and $\kappa > 0$, and let $f : Y \to X$ be a 1-Lipschitz function (Y dominates X, denoted as $X \prec Y$) then:*

1.
$$\mathrm{diam}(X, 1-\kappa) \leq \mathrm{diam}(Y, 1-\kappa)$$

2. $\mathrm{ObserDiam}(X; -\kappa) \leq \mathrm{diam}(X; 1-\kappa)$*, and* $\mathrm{ObserDiam}(X)$ *is finite.*

3.
$$\mathrm{ObserDiam}(X; -kappa) \leq \mathrm{ObserDiam}(Y; -kappa)$$

Proof. Since f is 1-Lipschitz, we have $f_*\mu Y = \mu_X$. Let A be any Borel set of Y with $\mu_Y(A) \geq 1-\kappa$ and $\overline{f(A)}$ be the closure of $f(A)$ in X. We have $\mu_X(\overline{f(A)}) = \mu_Y(f^{-1}(\overline{f(A)})) \geq \mu_Y(A) \geq 1-\kappa$ and by the 1-lipschitz property, $\operatorname{diam}(\overline{f(A)}) \leq \operatorname{diam}(A)$, so $\operatorname{diam}(X; 1-\kappa) \leq \operatorname{diam}(A) \leq \operatorname{diam}(Y; 1-\kappa)$.

Let $g : X \to \mathbb{R}$ be any 1-lipschitz function, since $(\mathbb{R}, |\cdot|, g_*\mu_X)$ is dominated by X, $\operatorname{diam}(\mathbb{R}; 1-\kappa) \leq \operatorname{diam}(X; 1-\kappa)$. Therefore, $\operatorname{ObserDiam}(X; -\kappa) \leq \operatorname{diam}(X; 1-\kappa)$.

and

$$\operatorname{diam}(g_*\mu_X; 1-\kappa) \leq \operatorname{diam}((f \circ g)_*\mu_Y; 1-\kappa) \leq \operatorname{ObserDiam}(Y; 1-\kappa)$$

□

References

[Axl23] S. Axler. *Linear Algebra Done Right*. Undergraduate Texts in Mathematics. Springer International Publishing, 2023. ISBN: 9783031410260. URL: https://books.google.com/books?id=0dnfEAAAQBAJ.

[BZ17] I. Bengtsson and K. Zyczkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2017.

[Fer] R. Feres. *Math 444 Lecture notes – the mathematics of quantum theory*. URL: https://www.math.wustl.edu/~feres/Math444Spring25/Math444Spring25Syllabus.html.

[GMS15] V. P. Gupta, P. Mandayam, and V. S. Sunder. *The Functional Analysis of Quantum Information Theory*. 2015. arXiv: 1410.7188 [quant-ph]. URL: https://arxiv.org/abs/1410.7188.

[Hay10] P. Hayden. "Concentration of measure effects in quantum information". In: *Quantum Information Science and Its Contributions to Mathematics*. Vol. 68. Proceedings of Symposia in Applied Mathematics. American Mathematical Society, 2010, pp. 211–260. ISBN: 978-0-8218-4828-9. DOI: 10.1090/psapm/068.

[HLW06] P. Hayden, D. W. Leung, and A. Winter. "Aspects of Generic Entanglement". In: *Communications in Mathematical Physics* 265.1 (Mar. 2006), pp. 95–117. ISSN: 1432-0916. DOI: 10.1007/s00220-006-1535-6. URL: http://dx.doi.org/10.1007/s00220-006-1535-6.

[KM] B. Kümmer and H. Maassen. "Elements of quantum probability". In: *Quantum Probability Communications*, pp. 73–100. DOI: 10.1142/9789812816054_0003. URL: https://www.worldscientific.com/doi/abs/10.1142/9789812816054_0003.

[Mec] E. Meckes. *The Random Matrix Theory of the Classical Compact Groups*.

[NC10] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[Pag] D. N. Page. "Page's conjecture". In: *Physical Review Letters* ().

[Par92] K. R. Parthasarathy. *An Introduction to Quantum Stochastic Calculus*. Vol. 85. Monographs in Mathematics. Birkhäuser Basel, 1992, pp. XI, 292. ISBN: 978-3-0348-9711-2. DOI: 10.1007/978-3-0348-8641-3.

[Par05] K. R. Parthasarathy. *Mathematical Foundation of Quantum Mechanics*. Vol. 85. Texts and Readings in Mathematics. Hindustan Book Agency, 2005, pp. XI, 292. ISBN: 978-93-86279-28-6. DOI: 10.1007/978-93-86279-28-6.

[San95] J. Sanchez-Ruiz. "Page's conjecture simple proof". In: *Physical Review E* (1995).

[Shi14] T. Shioya. *Metric measure geometry*. 2014. arXiv: 1410.0428 [math.MG]. URL: https://arxiv.org/abs/1410.0428.

[Ver18] R. Vershynin. *High-dimensional probability: an introduction with applications in data science*. Cambridge University Press, 2018.